Review on STEM OPT Policy and Impact

Henry Chen

July 17, 2023

Abstract

The effect matches to our hypothesis, that the 2016 STEM extension induces fewer immigrant-founded firms due to more rigidity in work. However, there are still some issues to be fixed before coming to a firmly conclusion.

Task

This report presents a preliminary analysis of the effects of local labor market exposure to the STEM Optional Practical Training (OPT) policy on variations in the number of both immigrant-founded and native-founded firms. Specifically, the study measures the share of STEM-related occupations at the city-year level to evaluate the local labor market's exposure to the STEM OPT policy. This macro exposure is inherently exogenous to an individual's micro choice of pursuing entrepreneurship or working in the private sector one or two years later. Consequently, this report investigates whether exposure to the STEM OPT policy could serve as an estimator of the exogenous supply shock in the immigrant entrepreneur landscape.

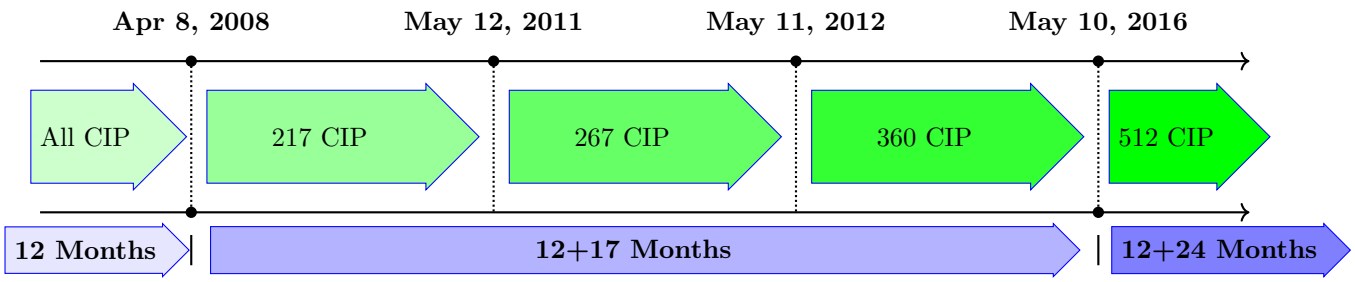
This report is structured as follows: First, a brief review of the context and key findings from previous analyses will be provided to set the stage for the current study. Building on these insights, the report posits a hypothesis regarding the impact of exposure to the STEM OPT policy on the immigrant and native entrepreneur market. Section 3 outlines the data construction process for measuring exposure and its integration with the existing dataset. Section 4 presents the estimation results, while Section 5 offers a conclusion that summarizes the report's key findings and implications.

Reviews and Hypothesis

STEM History

To start, this report provides a brief historical overview of the changes in the STEM OPT policy. Notably, the policy has undergone both expansions in the range of STEM majors and extensions in the duration of the program.

For the purpose of this analysis, we will concentrate on the significant OPT extension of eligible STEM majors that occurred in 2016.



Previous Analysis Results

In our previous analysis, we examined the trends in immigrant-founded entrepreneurship and native-founded firms, as well as the differences between the two groups, by regressing them on the year dummies corresponding to the STEM OPT policy changes. The initial hypothesis was that the extension of the STEM OPT policy would lead to an increase in immigrant-founded firms. This policy extension should provide more opportunities for immigrants to remain in the U.S. labor market and increase their chances of establishing their own businesses.

Table 1 and 2 present the regression results from the previous estimation. Table 1 focuses on the number of immigrant-founded firms after controlling for Foreign Direct Investment (FDI) and other sets of effects, while Table 2 demonstrates the difference-in-differences analysis, with immigrant status as the treatment variable. Surprisingly, both regression results indicate that the 2016 STEM OPT policy change led to a decrease in immigrant participation in the entrepreneurial market in subsequent years, which contradicts our initial hypothesis.

An alternative explanation for the impact of the STEM OPT policy change is that the extension allows immigrants to remain employed for longer periods, increasing their likelihood of continuing to work and reinforcing their attachment to the workforce. This could potentially result in less incentive for these individuals to establish their own businesses.

Table 1: Immigrant-Founded Firms

	(1) Level	(2) Level	(3) Growth	(4) Growth
Policy 2008	-0.796***		-0.668***	
	(-3.227)		(-3.142)	
Policy 2011	-0.287*		-0.005	
	(-1.762)		(-0.041)	
Policy 2012	0.154		0.176	
	(0.707)		(0.614)	
Policy 2016	0.834***		-0.483**	
	(3.059)		(-2.354)	
L.Policy 2008		-0.944**		-1.051***
		(-2.712)		(-3.644)
L.Policy 2011		0.128		0.542*
		(0.709)		(1.994)
L.Policy 2012		0.082		-0.631***
		(0.335)		(-3.552)
L.Policy 2016		0.749**		-0.818***
		(2.376)		(-3.051)
Fixed Effect	City	City	City	City
Cluster Std. Error	Cluster City+Year	Cluster City+Year	Cluster City+Year	Cluster City+Year
Obs	1598	1598	1120	1120

T-statistics in parenthesis

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 2: Firms founded by Treat/Control

	(1) ln	(2) ihs
Immigrant	-0.762***	-0.940***
	(-26.291)	(-29.009)
Treat X 2016 Dummy	-0.087***	-0.094***
	(-4.222)	(-3.784)
Controls		
Obs	24476	24476

T-statistics in parenthesis

* $p < .1$, ** $p < .05$, *** $p < .01$

New Hypothesis

Considering the unexpected findings from the previous analysis, we now anticipate that when focusing on STEM exposure, a higher exposure market after 2016 should lead to a decrease in the number of immigrant-founded firms in the market, as immigrant individuals become more firmly attached to their jobs. Therefore, the effect on the treated group—immigrants—should be expected to as a negative.

In specific, the coefficient on the $Treat_t \times Exposure \times Dummy_{2016}$ is negative.

Data Construction

Data

To carry out the analysis using the existing dataset, I have constructed a measure of occupational exposure to the STEM OPT policy by drawing upon several data sources. Specifically,

1. **Bureau of Labor Statistics (BLS)**

   The Occupation- and Industry-level employment by CBSA Code from 2002 to 2021 from BLS. The data records both the national employment data at each NAICS-SOC code, and SOC-level employment at each CBSA Code.

2. **National Center for Education Statistics (NCES)**

   The crosswalk between the major code CIP and the occupation code SOC. There are two crosswalks:

   - 2010 CIP to 2010 SOC
   - 2020 CIP to 2018 SOC.

3. **U.S. Immigration and Customs Enforcement (ICE) & Department of Homeland Security (DHS)**

   The list of STEM major codes at each STEM OPT policy amendment. There are four lists in total (i.e., 2008, 2011, 2012, and 2016)

4. **Office of Management and Budget (OMB)**

   List of Metropolitan Statistical Areas, as for reference of crosswalk city and MSAs in the data.

Exposure Measure Construction

The concept of local market exposure to the STEM OPT policy can be succinctly expressed as the employment-weighted-average share of STEM-related occupations, represented as a number between 0 and 1. A higher value for this exposure measurement signifies that a city (or Metropolitan Statistical Area, MSA) has a greater proportion of its occupations likely to be filled by graduates from STEM programs or majors.

The likelihood of an occupation being filled by a STEM graduate is determined by calculating the ratio of the number of STEM majors related to that job to the total number of majors associated with it. Then the exposure is the average of the occupation's STEM share, weighted by the number of employment in this occupation in that city (or MSA).

To express it mathematically, the exposure measured, $Exposure_{t,i}$, for city i in year t

$$Exposure_{t,i} = \sum_{j}^{N} \left(w_j \times \frac{\sum_{q}^{STEM_j} \mathbb{1}}{\sum_{p}^{M_j} \mathbb{1}} \right)$$

where N denotes the set of occupations in i at year t; w_j is the share of employment in this occupation relative to the total employment in i at year t; $STEM_j$ is the set of STEM majors related to this occupation j, and M_j is the set of majors related to this occupation j. Each set will vary across years.

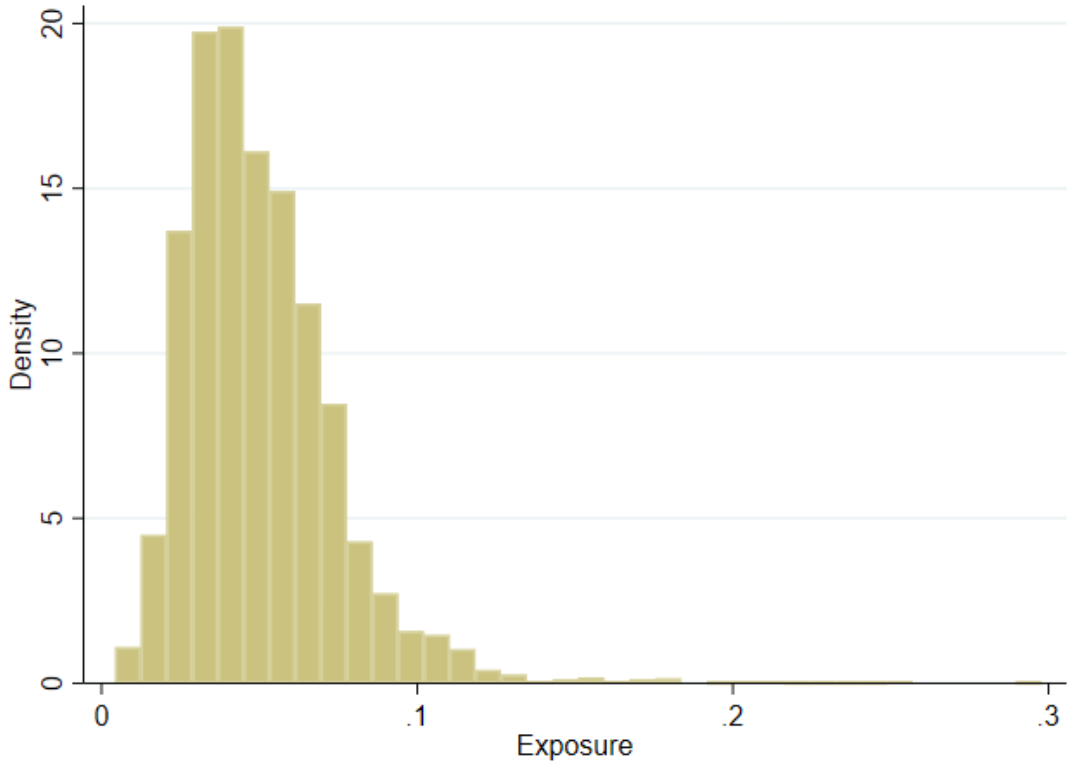


Figure 1: Exposure Distribution

Summary Statistics for Exposure Measurement

The exposure measurement spans from 2011 to 2021, based on the availability of occupation employment data at the Core-Based Statistical Area (CBSA) level. A summary of statistics for the exposure measurement is provided in Table 3. With a mean of 0.05, this indicates that, on average, 5 percent of occupations are filled by STEM graduates.

Figure 1 displays the distribution of the exposure measurement. As the histogram illustrates, most statistical areas have low exposure to the STEM policy, as STEM major is only a niche branch in all majors. A more detailed breakdown, differentiating between MSAs and non-MSAs, is depicted in Figure 2.

Table 3: Summary Statistics

	Obs	Mean	Std	Min	P25	Median	P75	Max
Exposure	4286	0.0505	0.0248	0.00430	0.0335	0.0465	0.0626	0.298

Upon examining the distribution for the year 2016, as illustrated in Figure 3, we observe that more areas are exposed to STEM occupations (at lower/medium tail), but there are fewer areas with extreme exposure (at higher tail). It is important and interesting to note that the mean exposure in the **post-2016** sample is lower than the mean in the **pre-2016** sample (0.048 compared to 0.053). This change is likely due to the crosswalk between major codes and occupation codes, as there are only two available sources throughout the sample. I attempted to use only the 2010 crosswalk for the sake of consistency, but the National Center for Education Statistics (NCES) periodically modifies occupation codes, resulting in an increased number of mismatches when relying on a single crosswalk.

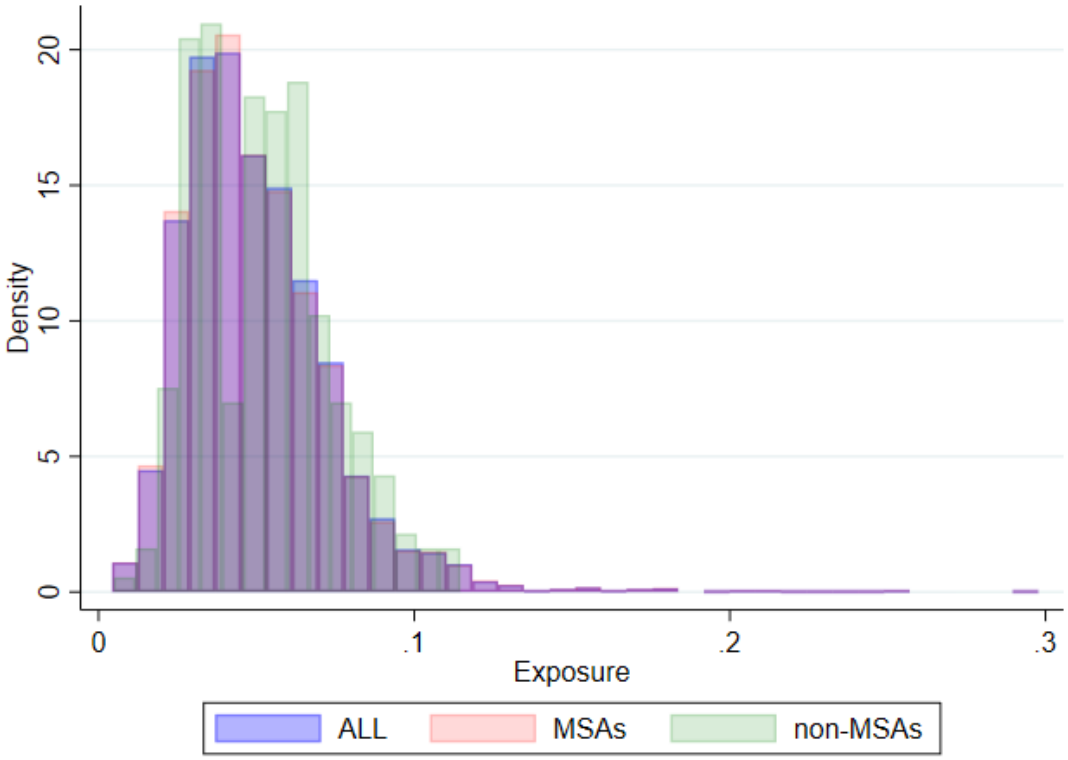


Figure 2: Exposure Distribution: MSA/non-MSA

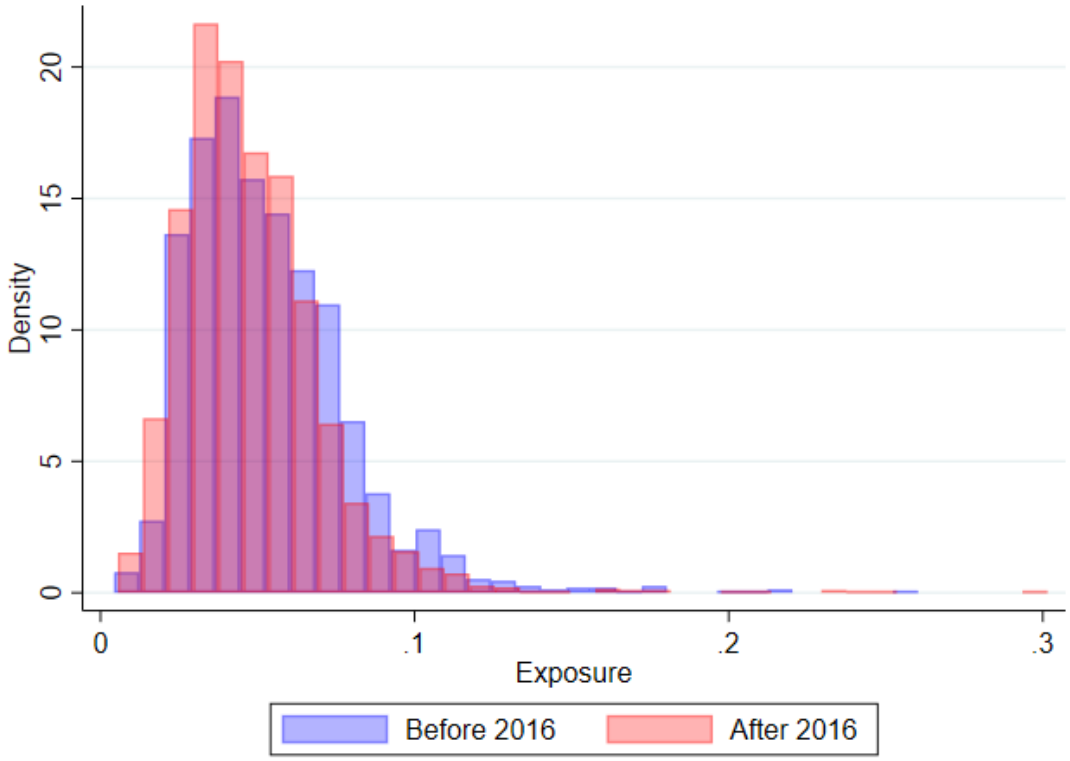


Figure 3: Exposure Distribution: Pre-/Post-2016

Regression Results

Appending from the previous sample, I add the city-level exposure to the STEM OPT policy, which is proxied by the city-corresponding MSA's exposure. Upon the previous regression model in Table 2, I further regress on exposure measurement, such that

$$\begin{aligned}\ln(\#\text{ of Firms founded})_{i,t} =& \beta_1 * \mathbf{1}(\text{Immigrant})_{i,t} \\ &+ \beta_2 * (\text{STEM OPT exposure})_{i,t} \\ &+ \beta_3 * \mathbf{1}(\text{Year} \geq 2016)_{i,t} \\ &+ \beta_4 * \mathbf{1}(\text{Immigrant})_{i,t} * \mathbf{1}(\text{Year} \geq 2016)_{i,t} \\ &+ \beta_5 * \mathbf{1}(\text{Immigrant})_{i,t} * (\text{STEM OPT exposure})_{i,t} \\ &+ \beta_6 * \mathbf{1}(\text{Immigrant})_{i,t} * (\text{STEM OPT exposure})_{i,t} * \mathbf{1}(\text{Year} \geq 2016)_{i,t} \\ &+ Control + \epsilon_{n,t}\end{aligned}$$

whereas controlling the year- and city-fixed effect, weighted by the natural log of the number of firms, and clustered standard error at city and year.

Table 4 reports the estimation results

In line with the previous analysis results, the coefficient for the Treatment Group is negative, indicating that native-founded firms outnumber immigrant-founded firms. The variable "Exposure" has a positive coefficient, suggesting that a more STEM-related labor market tends to foster a greater number of immigrant-founded

Table 4: Append Exposure

	(1) Log	(2) IHS
Treat: Immig.	-1.183***	-1.452***
	(0.0910)	(0.100)
Exposure	-1.608*	-2.059**
	(0.795)	(0.904)
Treat X 2016 Dummy	0.0376	0.0568
	(0.0666)	(0.0789)
Treat X Exposure	3.912***	4.947***
	(1.021)	(1.122)
Treat X Exposure X 2016 Dummy	-0.281	-0.415
	(0.437)	(0.525)
R-square	0.898	0.883
Obs	12594	12594

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

firms. The interaction term involving the treatment, 2016 dummy, and exposure is negative, implying that a higher STEM exposure market is less favorable for immigrant entrepreneurs. This may reduce their inclination to start businesses and increase their attachment to their jobs. However, the coefficient is insignificant, which could be attributed to the sample size issue arising from matching BLS MSAs to the cities in our data.

In another attempt, where I use all MSA crosswalks throughout history, the coefficient becomes more significant even though the increase in matching is relatively small. This outcome is observed when compared to the results obtained using only the 2017 crosswalk. Table 5 presents these findings.

Current Issues

There are two primary issues that I am currently addressing in this analysis report. The first one concerns the crosswalk. When merging the exposure measurement calculated from the Bureau of Labor Statistics (BLS) onto our dataset, there are some mismatches between the MSAs from BLS and the MSAs from city crosswalks in our data. I am in the process of refining the crosswalk between the BLS data and our dataset's city-MSA crosswalk.

The second issue I encountered arises when adding the "Exposure x 2016 Dummy" term to the regression. The coefficient for the triple interaction term becomes significant in this case. At present, I have not formulated an explanation for the inclusion of this term or how to interpret this coefficient. Table 6 and 7 display the estimation results for this scenario.

Table 5: Append Exposure (More Crosswalk)

	(1) Log	(2) IHS
Treat: Immig.	-1.165***	-1.430***
	(0.0869)	(0.0971)
Exposure	-1.533*	-1.941**
	(0.690)	(0.785)
Treat X 2016 Dummy	0.0454	0.0684
	(0.0604)	(0.0718)
Treat X Exposure	3.776***	4.756***
	(0.969)	(1.072)
Treat X Exposure X 2016 Dummy	-0.362	-0.540
	(0.373)	(0.437)
R-square	0.897	0.882
Obs	13726	13726

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 6: Append Exposure

	(1) Log	(2) IHS
Treat: Immig.	-1.205***	-1.479***
	(0.0853)	(0.0954)
Exposure	-1.744*	-2.221**
	(0.830)	(0.949)
Treat X 2016 Dummy	0.0769	0.103
	(0.0538)	(0.0693)
Treat X Exposure	4.169***	5.253***
	(0.932)	(1.035)
Exposure X 2016 Dummy	0.599	0.712
	(0.424)	(0.490)
Treat X Exposure X 2016 Dummy	-0.775**	-1.002**
	(0.324)	(0.381)
R-square	0.898	0.883
Obs	12594	12594

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 7: Append Exposure (More Crosswalk)

	(1) Log	(2) IHS
Treat: Immig.	-1.192***	-1.462***
	(0.0818)	(0.0928)
Exposure	-1.699**	-2.139**
	(0.714)	(0.816)
Treat X 2016 Dummy	0.0909*	0.123*
	(0.0477)	(0.0630)
Treat X Exposure	4.081***	5.122***
	(0.883)	(0.987)
Exposure X 2016 Dummy	0.692	0.829*
	(0.385)	(0.454)
Treat X Exposure X 2016 Dummy	-0.939***	-1.229***
	(0.243)	(0.282)
R-square	0.897	0.882
Obs	13726	13726

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$